DOCUMENT RESUME

ED 074 155 TM 002 520

AUTHOR Hofmann, Richard J.
TITLE Indices of Complexity and Interpretation: Their Computation and Uses in Factor Analysis.
PUB DATE Mar 73
NOTE 16p.; Paper presented at annual meeting of the National Council of Measurement in Education (New Orleans, Louisiana, March 1973)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Cluster Analysis; *Factor Analysis; *Factor Structure; Speeches; Statistical Analysis; Tables (Data); Technical Reports

ABSTRACT

In this methodological paper two indices are developed: a complexity index and an interpretation index. The complexity index is a positive number indicating on the average how many factors are used to explain each variable in a factor solution. The interpretation index will be positive ranging from zero to unity; unity representing a perfect independent cluster solution and zero representing the poorest factor solution in terms of complexity. Through empirical application to the classic 24 psychological variables it is demonstrated that the indices may be computed by hand and are easily interpreted providing a basis for comparing different factor solutions. (Author)

27.18

Indices of Complexity and Interpretation: Their Computation and Uses in Factor Analysis

Richard J. Hofmann
Miami University

Paper presented at annual meeting of
The National Council of Measurement in Education,
New Orleans, March 1973

Usually when factor analysts refer to simple structure in a factor solution their reference is based upon what might typically be referred to as a perfect independent cluster solution, all loadings in the factor matrix are either zero or unity. Although the literature abounds with indices of simple structure, these indices are primarily analytic computational criteria for orthogonal and oblique transformation solutions and have little apparent meaning to the field practitioner. For the practitioner there is no easily interpreted index of simple structure; thus many users of factor analysis are forced to accept a solution on faith.

The objective of this manuscript is the algebraic and logical development of several new indices that are somewhat related to simple structure indices. These indices are descriptive of the complexity and general interpretability of a factor solution and they will be computed as a function of the squared and quartic powers of the entries of a row normalized matrix.

Complexity by Row: Variable Complexity

In this section the concern will be with variable complexity. The complexity of a variable is the number of non-vanishing entries in its associated row of a factor matrix. Specifically it is the number of factors describing a variable in a particular factor solution. Thurstone (1947) indicates that one of the objectives of factor analysis is finding the smallest number of factors for describing each variable and that variables which are of low complexity are good indices of the nature of a factor.

Assume that n variables are defined by r factors in the (n x r) factor matrix F. Regardless of whether F is an orthogonal or oblique solution matrix compute the row sums of squares and let the sum of the squares of the i^{th} row be denoted as the ii^{th} entry of the diagonal matrix H^2. Premultiplication of F by H^{-1}, the inverse of the square roots of the row sums of squares, will result in a matrix, V, whose row sums of squares are unity.

$$H^2 = \text{diagonal } (FF'); \quad (1)$$

$$V = H^{-1}F. \quad (2)$$

As the complexity of any row of F and hence its row normalized analog, V, is reduced the proportionate contribution of the largest row entry to the row sum of squares is increased. Regardless of the magnitude of a row entry of F as the row tends toward complexity one the major entry of its row normalized analog, V, tends toward unity. For perfect unit complexity, the average variable complexity being unity, in the matrix F the entries of all rows of the normalized matrix V will be either zero or unity.

Assume that the complexity of some row i of F is a. That is, a factors define the variable i. The row sum of squares for variable i in the normalized matrix, V, is unity, however it may be referred to as (a/a). If each of the a factors defining i makes an equal contribution to the row sum of squares for i then each entry of the i^{th} row of V will be either zero or ($1/a^{\frac{1}{2}}$) inasmuch as the following equality

$$1 = \underline{a}/\underline{a} = (1/\underline{a}^{\frac{1}{2}})_1^2 + (1/\underline{a}^{\frac{1}{2}})_2^2 + (1/\underline{a}^{\frac{1}{2}})_3^2 + \ldots + (1/\underline{a}^{\frac{1}{2}})_{\underline{a}}^2 \qquad (3)$$

must hold for a normalized row of complexity $\underline{a}$. Although any row sum of squares will always be equal to unity for the normalized matrix, $\underline{V}$, the row sum of the quartic loadings will only be unity when the row is of complexity one. The quartic loadings of a normalized row of complexity $\underline{a}$ will be either zero or $(1/\underline{a}^{\frac{1}{2}})^2$. The sum of the quartic loadings of a normalized row of complexity $\underline{a}$ is

$$1/\underline{a} = \underline{a}/\underline{a}^2 = (1/\underline{a})_1^2 + (1/\underline{a})_2^2 + (1/\underline{a})_3^2 + \ldots + (1/\underline{a})_{\underline{a}}^2 \qquad (4)$$

just the reciprocal of the complexity of the variable associated with the row. The reciprocal of the row sum of the quartic loadings of a normalized row is analagous to the complexity of the variable associated with the row.

In practice the complexity of a single variable will seldom be a whole number, however when applied to all variables an average complexity although not a whole number will still be meaningful. An index of average variable complexity, $\underline{c}_1$, may be computed by summing the complexity of each individual variable and then dividing by the total number of variables. Define average complextiy as

$$\underline{c}_1 = 1/\underline{n} \sum_{i=1}^{n} \left(\frac{1}{\sum_{j=1}^{r} v_{ij}^4} \right) \qquad (5)$$

where the value $\underline{v}_{ij}$ is just the $\underline{ij}^{th}$ entry of the row normalized matrix $\underline{V}$.

Interpreting the Index of Row Complexity

The most interpretable complexity for any row is unity. For a total matrix the most interpretable, but not necessarily the most compelling, complexity would be unity. Inasmuch as departure from unit complexity is only in one direction unity is the minimum value attainable for $\underline{c}_1$. The poorer the complexity of a factor solution the more marked will be the departure of $\underline{c}_1$ from unity with a maximum value for the complexity index being $\underline{r}$ the total number of factors. For any given set of data the average complexity will be some value between unity and $\underline{r}$, the number of factors, inclusive with a range of $(\underline{r} - 1)$.

The complexity index is then a function of the number of factors. Although it is clear that a complexity index of unity is the lowest possible index and a complexity index of $\underline{r}$ is the largest possible index it is not clear how one might evaluate a complexity index that is not an extreme value. It is possible to develop an index of relative interpretation to be considered with the complexity index.

Compute the difference between maximum complexity and observed complexity, $(\underline{r} - \underline{c}_1)$, and express that difference as a ratio to, $(\underline{r} - 1)$, the difference between maximum and minimum complexity. This index will always be a positive number ranging between zero and unity. For the poorest possible row complexity the index of row interpretability will be zero and for the best possible row complexity the index of row interpretability will be unity. Define the index of row interpretability as:

$$\underline{c}_2 = \frac{\underline{r} - \underline{c}_1}{\underline{r} - 1} \qquad (6)$$

Complexity by Column: Factor Complexity

Kaiser (1958) suggests that a most fundamental requirement for simple structure may be one of simplifying the columns of a factor matrix. Indeed his varimax criterion when maximized will lead, according to him, to the maximum interpretability or simplicity for an entire factor matrix. However, maximum interpretability or simplicity of the columns of a factor matrix is discussed in the literature from at least two points of view.

One point of view is typified by Kaiser's (1958) attempts to maximize the variance of the squared column loadings of an orthogonal factor solution. Alternatively Saunders' (1962) equamax in tending to equalize the contributions of each factor actually tends to minimize the variance of the squared column loadings of an orthogonal factor solution.

The index of variable complexity may be unity for either a Kaiser-type solution or a Saunders-type solution. The index yields no information pertaining to the contributions of the factors to the variable complexity. That is, it is not possible to tell whether the factors make level contributions to the variables or whether there is a tendency for a single factor to account for the variable complexities, especially if the index of variable complexity is small. What might be said about the complexities of the factors?

Unlike the variable complexity the average number of variables described by a factor will always be known inasmuch as there will always be n variables and r factors; thus an average factor complexity of (n/r). Assuming that the matrix has been row normalized, as described in equations 1 and 2, it is possible to make inferences about column complexity indirectly by observing the squared deviations of the column sums of squares from (n/r).

Assume an extreme case of a general factor and $(r - 1)$ imaginary factors. In this particular case the sum of squares would be zero for each column associated with an imaginary factor. For the one general factor the column sum of squares would be n. This extreme case defines the situation that must exist in order to obtain the maximum sum of the squared deviations of the column sums of squares from (n/r). This maximum sum is just $(n^2/r)(r-1)$.

As the variable loadings become more evenly distributed across all r columns or factors the sum of the squared deviations of the column sums of squares from (n/r) will tend toward zero. If every loading in the factor matrix F is identical to every other loading then every loading in the row normalized matrix, V, will be equal to the reciprocal square root of the number of factors, $(1/r)^{\frac{1}{2}}$, and the sum of the squared deviations of the column sums of squares from (n/r) will be exactly zero. If however every variable is of complexity one, and if each column of F has (n/r) loadings in it (assuming n is perfectly divisible by r)

then the sum of the squared deviations of the column sums of squares from (n/r) computed from V, will again be zero. Thus for two extreme types of solutions a general factor and each variable having a complexity of r, the sum of the squared deviations of the columns sums of squares from (n/r) will be zero when computed from the row normalized matrix. The following inequality is indicative of the upper and lower bounds for the observed sum of the squared deviations when they are computed according to the middle term of the inequality.

$$0 \leq \sum_{j=1}^{r} \left(\sum_{i=1}^{n} v_{ij} - n/r \right)^2 \leq (n^2/r)(r-1) \qquad (7)$$

where v_{ij} is the value in the j^{th} column of the i^{th} row of the row normalized matrix V. The right term in the expression is the maximum sum of squares.

The ratio of observed sum of squares to maximum sum of squares will always be some value between zero and unity. When the ratio is zero the simplicity of the columns of a factor matrix will be ideal in the sense that each factor will have a complexity of (n/r). When the ratio is unity the factor matrix will be "complex" from the point of view that there will be just one factor accounting for all n variables. If the ratio is subtracted from unity an interpretation may be made that is similar to the interpretation accorded to the variable complexity. Define column interpretability, c_3, as;

$$c_3 = 1 - \frac{\sum_{j=1}^{r} \left(\sum_{i=1}^{n} v_{ij}^2 - \frac{n}{r} \right)^2}{\left(\frac{n^2}{r}\right)(r-1)} \qquad (8)$$

and then for what many factor analyst would refer to as undesirable factor complexity, a general factor, a value of zero will result while for the type factor solution found desirable by many factor analyst, level factor complexity, a value of unity will result.

Complexity Equilibration:

A Balance Between Row and Column Complexity

Regretably the index of either row or column interpretability can be high for a solution that has some sort of undesirable complexity. If both indices are high for a given solution then the solution should be desirable in the sense that the variable complexity will be low and the factor complexity will be neither high nor low. There seems to be no apparent situation in which both indices are low. It would seem reasonable to conclude that the interpretation indices both converge toward unity when a solution has the most desirable complexity: each row being of unit complexity and each column being of complexity (n/r). This "equilibrium" between row and column interpretability would seem to be a most desirable property for a simple structure-type factor solution.

In previous sections of this manuscript row interpretability and column interpretability were discussed independently of each other. It was noted that both indices will range from zero to unity. With a bit of algebraic manipulation it may be demonstrated that row interpretability varies as a function of $\underline{n} \sum_{i=1}^{n} (1/ \sum_{j=1}^{r} v_{ij}^{4})$ while column interpretability varies as a function of the term $\underline{r}\sum_{j=1}^{r}(\sum_{i=1}^{n}\underline{v}_{ij}^{2})^{2}$ and both terms have precisely the same limits

$$\underline{n}^2 \leq \underline{r} \sum_{j=1}^{r}(\sum_{i=1}^{n} \underline{v}_{ij}^2)^2 \leq \underline{n}^2\underline{r} ;$$

$$\underline{n}^2 \leq \underline{n} \sum_{i=1}^{n} (1/ \sum_{j=1}^{r} \underline{v}_{ij}^4) \leq \underline{n}^2\underline{r} .$$

The lower bound of the inequality is associated with a maximum interpretation index while the upper bound is associated with a minimum interpretation index. Complexity equilibrium may be expressed as the difference between the two middle terms in the above inequalities.

$$\underline{r} \sum_{j=1}^{r} (\sum_{i=1}^{n} \underline{v}_{ij}^2)^2 - \underline{n} \sum_{i=1}^{n} (1/ \sum_{j=1}^{r} \underline{v}_{ij}^4) \qquad (9)$$

In the previous sections discussion centered about two types of undesirable solutions and the interpretation indices were developed around these undesirable solutions. The first type of undesirable solution is characterized by the solution in which every value in the row normalized matrix is equal to $1/\underline{r}^{\frac{1}{2}}$, every variable is of complexity $\underline{r}$. For this solution column interpretability will be a maximum while row interpretability will be a minimum and

$$\underline{r} \sum_{j=1}^{r} (\sum_{i=1}^{n} \underline{v}_{ij}^2)^2 - \underline{n} \sum_{i=1}^{n} (1/ \sum_{j=1}^{r} \underline{v}_{ij}^4) = -\underline{n}^2(\underline{r}-1).$$

The second type of undesirable solution is characterized by a general factor and ($\underline{r}$-1) imaginary factors. For this solution column interpretability will be a minimum while row interpretability will be a maximum and

$$\underline{r} \sum_{j=1}^{r} (\sum_{i=1}^{n} \underline{v}_{ij}^2)^2 - \underline{n} \sum_{i=1}^{n} (1/ \sum_{j=1}^{r} \underline{v}_{ij}^4) = \underline{n}^2(\underline{r}-1).$$

Assuming that a general factor solution is more desirable than a solution in which all the variables are of complexity $\underline{r}$ then expression 9 will approximate zero when there is an equilibrium between row and column complexity and it will be negative when the row complexity is not as interpretable as the column complexity and positive when the column complexity is not as interpretable as the row complexity. Utilizing the maximum difference that may be achieved for expression 9 it is possible to derive an index of equilibration, $\underline{c}_4$:

$$\underline{c}_4 = \frac{\underline{r}\sum_{j=1}^{r}\left(\sum_{i=1}^{n} \underline{v}_{ij}^{2}\right) - \underline{n}\sum_{i=1}^{n}\left(1/\sum_{j=1}^{r} \underline{v}_{ij}^{4}\right)}{\underline{n}^{2}\,(\underline{r}-1)} . \tag{10}$$

The values of the equilibration index will range between positive and negative unity. Actually the index is a measure of disequilibruim as a value of either positive or negative unity will be indicative of total disequilibrium. For a value of negative unity every variable will be of complexity $\underline{r}$ while every factor will define $\underline{n}/\underline{r}$ variables. For a value of positive unity every variable will be of complexity one but there will be one general factor and ($\underline{r}$-1) imaginary factors.

When the equilibration index is zero the column and row interpretations, or complexities, are in perfect equilibrium. The equilibration index may achieve its most desirable value, zero, even when the variables are not of unit complexity. There are some types of solutions in which unit complexities for certain variables is undesirable, such as the Thurstone (1947) box problem. For such solutions there is a certain implicit restriction placed on the magnitudes of the interpretation indices, however there can still be equilibration between rows and columns. As the interpretation indices converge toward each other the equilibration index will tend toward zero, thereby not being restricted by the absolute magnitudes of the interpretation indices.

Empirical Application

In order to demonstrate the applicability of the indices developed in this manuscript they were computed from a variety of orthogonal and oblique transformation solutions of the Holzinger and Harman (1941) centroid solution of the 24 psychological variables.

Following Saunders (1962) twelve orthogonal transformation solutions were generated, six normal and six raw solutions. The orthomax weight was varied in the orthomax criterion from zero to two and one-half in increments of one-half. The solutions resulting from these variations included the quartimax, varimax and equamax transformations. In addition two oblique primary pattern matrices were taken from Harman (1967). These two oblique solutions, the oblimax and the biquartimin, represent reasonably good solutions for the data.

Each solution matrix was row normalized and the indices of complexity and interpretation were computed. These indices are reported in Table 1.

<u>Table 1 about here</u>

There are several important generalizations which may

Table 1 Indices of Complexity and Interpretation

Applied to Transformations of the Centroid Solution

of the Twenty-four Psychological Variables

Type Solution	Orthomax Weight	Row Complexity	Interpretation Row	Interpretation Column	Equilibration
<u>Raw</u>					
Quartimax	0.0	1.97	.68	.94	-.26
-	0.5	1.93	.69	.98	-.29
Varimax	1.0	1.90	.70	.99	-.29
-	1.5	1.89	.70	.99	-.29
Equamax	2.0	1.88	.71	1.00	-.29
-	2.5	1.88	.71	1.00	-.29
<u>Normal</u>					
Quartimax	0.0	1.72	.76	.80	-.04
-	0.5	1.82	.73	.98	-.25
Varimax	1.0	1.85	.72	.99	-.28
-	1.5	1.87	.71	1.00	-.28
Equamax	2.0	1.87	.71	1.00	-.29
-	2.5	1.88	.71	1.00	-.29
Oblimax	--	1.51	.83	.98	-.15
Biquartimin	--	1.53	.82	.99	-.16
Centroid Solution	--	2.00	.67	.67	.00

be made from Table 1. Kaiser (1958) in comparing the varimax solution with the equamax solution of the 24 psychological variables noted that the quartimax had a tendency to load highly on the first factor, the dominant factor, and have non zero loadings on subsequent factors. Such a solution should be characterized by a rather large complexity index, a relatively low row interpretation index, and a relatively high equilibration index. Note however that the normal quartimax has a relatively low index of complexity, at least for the orthogonal solutions, an exceptionally low equilibration index, but also a very low index of column interpretability. Such a low index of column interpretability suggests that the solution is based primarily on a general factor. This may be verified by looking at the normal quartimax solution in Table 2.

Also worthy of mention was the tendency for the variable complexity to become smaller as the orthomax weight was increased for the raw transformation solutions. The decrease in variable complexity was only temporary as the index stabilized at a minimum value of 1.88. Alternatively for the normal transformation solutions the complexity index increased as the orthomax weight increased. This trend was also only temporary as the index stabilized at a maximum value of 1.88. It would seem as though 1.88 may represent the best attainable simple structure for an orthogonal solution even though it does not represent the lowest attained index of complexity.

<u>Table 2 about here</u>

The oblique solutions showed considerably less variable

Table 2 Centroid Solution and Various Transformations of the Centroid Solution of the Twenty-four Psychological Variables*

	Centroid				Transformation Solution															
					Raw Equamax				Normal Quartimax				Normal Varimax				Oblimax			
1	61	-12	30	-25	19	17	66	20	72	03	-10	-07	14	19	67	17	-13	02	83	-05
2	37	-12	21	-14	13	10	42	08	46	-03	-05	-05	10	07	43	10	-06	-06	55	-03
3	43	-22	26	-16	19	09	53	03	55	-10	-03	-10	15	02	54	08	-02	-15	70	-09
4	48	-21	21	-18	23	08	53	10	58	-04	01	-11	20	09	55	07	03	-06	68	-13
5	67	-31	-34	11	76	10	16	23	51	14	62	03	75	21	22	13	84	10	-04	-09
6	66	-34	-26	22	77	19	17	12	52	02	62	11	75	10	23	21	83	-07	-02	05
7	65	-40	-38	12	83	05	15	18	50	09	70	-01	82	16	21	08	95	05	-05	-16
8	66	-23	-15	-06	56	10	34	27	59	15	38	-03	54	26	38	12	51	15	26	-13
9	66	-39	-24	31	82	23	16	03	52	-07	67	16	80	01	22	25	91	-19	-05	12
10	46	46	-37	-14	14	19	-08	71	21	69	11	21	15	70	-06	24	07	83	-38	12
11	57	40	-21	-06	18	32	06	62	36	56	09	28	17	60	08	36	03	64	-20	27
12	48	37	-15	-39	03	07	22	69	38	62	-09	-01	02	69	23	11	-18	79	14	-12
13	61	13	-10	-40	20	03	40	60	56	48	01	-11	18	60	41	06	-02	62	39	-24
14	44	20	-01	29	23	48	01	19	29	13	15	45	22	16	04	50	14	04	-22	59
15	41	17	15	27	14	49	11	10	33	02	02	42	12	07	14	50	-02	-09	-02	61
16	52	08	30	08	12	44	39	12	54	-01	-09	27	09	10	41	43	-15	-10	40	43
17	49	32	08	34	16	62	03	22	33	15	05	57	14	18	06	64	-01	04	-21	79
18	55	31	25	07	03	53	30	29	50	18	-16	40	00	27	32	54	-28	12	22	58
19	45	13	13	11	15	39	22	18	40	08	01	29	13	15	24	39	-03	20	14	40
20	61	-17	13	00	38	25	44	12	63	-02	16	08	35	11	48	25	21	-09	46	11
21	60	11	08	-17	18	24	40	40	57	27	-02	09	15	38	42	26	-07	30	39	10
22	61	-14	15	14	40	36	38	07	60	-07	18	20	36	05	41	36	23	-18	35	29
23	69	-16	13	-12	39	21	54	22	72	06	13	01	35	21	57	22	17	03	59	00
24	65	15	-15	00	36	31	19	46	50	36	21	22	34	44	22	34	23	38	-01	21

*All loadings have been multiplied by 100 to eliminate decimals.

complexity than did the orthogonal solutions. Because of this the row interpretations were larger than they were for the orthogonal solutions and the column interpretations remained about the same. Consequently the index of equilibration has been reduced for the oblique solutions. Within the framework of simple structure one may conclude that the oblique solutions are better than the orthogonal solutions.

Although it may be artifactual both row and column interpretability are identical for the centroid solution. Additionally these interpretation indices are the lowest of all computed and the complexity is the largest computed index. Clearly then one should realize that the associated equilibration index of zero is not indicative of the best simple structure solution.

Unfortunately the index of row complexity may be low yet the associated solution matrix may not be easily interpreted. Such a situation occurs with the normal quartimax solution reported in Table 2.

Finally we will interpret the indices as being descriptive of the oblimax solution.

(a) The average number of factors describing each variable in the oblimax solution is 1.51.

(b) Considering the poorest complexity possible for this matrix the observed solution is 83 per-cent efficient. Such an index indicates that a complexity of 1.51 is not a large complexity for this data set.

(c) The index of column interpretation of .98 is indicative of any approximate equal spread of the factor loadings across all four factors. That is, there is no tendency toward a general factor.

(d) The equilibration index being -.15 indicates that the factors of the matrix are less complex, relative to their maximum possible complexity, than are the variables, relative to their maximum possible complexity.

Conclusion

Although this manuscript was initially conceived of in terms of row complexity, three additional indices were also developed. These additional indices were developed in order to obviate certain problems associated with the complexity index.

These indices are descriptive of column and row complexities in a factor matrix, regardless of whether the solution is orthogonal or oblique. Although these indices bear some computational resemblance to numerous simple structure criteria no one index will serve as a satisfactory simple structure transformation criterion. Although not explicitly reported in this manuscript we have attempted to combine these indices into a simple structure criterion with a total lack of success.

The single major virtue of these indices is that they are descriptive of a factor solution, and are easily understood whether applied to an orthogonal or oblique solution matrix..

References

Harman, H. Modern factor analysis. (Second Edition). Chicago: University of Chicago Press, 1967.

Holzinger, K. and Harman, H. Factor analysis a synthesis of factorial methods. Chicago: University of Chicago Press, 1941.

Kaiser, H. The varimax criterion for analytic rotation in factor analysis. Psychometrika, 1958, 28, 187-200.

Saunders, D. Trans-varimax: Some properties of the ratiomax and equamax criteria for blind orthogonal rotation. Paper read at annual meeting of American Psychological Association. St. Louis, 1962.

Thurstone, L. Multiple factor analysis. (seventh impression). Chicago: University of Chicago Press, 1947.